Experimental Evidence on Financial Incentives, Information and Decision Making†

by

Ian M. Dobbs

and

Anthony D. Miller

Newcastle University Business School
Newcastle upon Tyne
NE1 7RU

JEL Classification: C91, D83, J33, M52.

Keywords: Experiments, Incentives, Learning.

†The authors are grateful for the financial support of the Institute of the Chartered Accountants of Scotland (Project Nos. RES0371 07208 099 and RES0371 07209).

We also thank Michelle Crickett, Christine Helliar, John Innes, and the participants at conferences and workshops at the Universities of Heriot Watt, Northumbria, Royal Holloway and Wollongong, for helpful comments. Finally, our thanks to the editor and two anonymous referees whose comments have helped us to materially improve the paper.

ABSTRACT

This paper reports on a series of experiments aimed at exploring in greater detail previous work on the effects of financial incentives on information use and task performance in a principal-agent setting. The paper also develops some new statistical modelling in the area of experimental testing including incorporation into the modelling approach of data from post-experiment questionnaires. We find significant support for the finding that profit-related individual money rewards encourage increased accessing of valuable but costly past profit information by agents - and that this in turn enhances individual performance in earning profit for the agency.

1. Introduction

This paper makes use of a research design based on Sprinkle (2000) in order to re-examine the previous finding that, compared to fixed rewards, performance-related rewards increase individual demand for valuable information and enhance performance. This issue is of major importance to the design of performance measurement and reward systems based on accounting information, since agency theory suggests that performance-related rewards are necessary to mitigate moral hazard and adverse selection problems in complex organisations. Our study extends the earlier work in that the level of rewards is systematically varied over a much wider range (in the original study only two levels were considered), and the sample size is also significantly increased. Our sample also has rather different demographics compared to the original work; instead of featuring exclusively US sophomore participants, our sample featured final year undergraduate and postgraduate students with a non-US international background (primarily European and Asian students). This increased diversity within the sample allowed us to explore additional research questions - in particular, concerning the role of GPA, gender, age and nationality in explaining task performance. As a final extension, we implemented a post experiment questionnaire and used the data from this as an integral part of the statistical testing procedure. Whilst post-experiment questionnaires have been used in previous work, they have been used exclusively to eliminate subjects who reveal some apparent imperfections in their understanding (make errors in one or more of their answers to the post experiment questions). We take the view that responses to such questionnaires, just as with responses within the experiment, are properly viewed as stochastic variables, and that it is therefore appropriate to explicitly model this within the overall analysis of the experiment as a whole.

Information is conventionally considered as offering two distinct potential benefits to organisations. First, it can inform a decision-maker about significant factors affecting *future* performance that should be taken into account in reaching a decision. This function may be termed its *decision-facilitating* role. For example, demand information would be expected to affect a company's stock-ordering and holding policies (Feltham, 1972), and cost variances might be used to determine whether the current state of production equipment is satisfactory (Chow et al, 1990). Second, information can provide measures of *past* performance that, if linked to extrinsic employee rewards, might motivate employees to try to perform well in order to gain favourable ex post evaluations of performance and, in turn, higher rewards; see Baiman (1982, 1990) for applications to managerial accounting principles and practices such as responsibility accounting, cost allocation and budgeting.

This function may be termed its *decision-influencing* role. When an environment is stable, so that past periods are related to present and future periods, then *feedback* information on past performance can potentially provide both decision-facilitating and decision-influencing benefits. This is typically the case for accounting information; see for example Baiman & Demski (1980) on conditional variance investigation policies.

In laboratory studies, empirical corroboration of the importance of these two roles for information has, until recently, focussed mainly on their separate investigation. Researchers have extensively studied how decision-makers use information in reaching their decisions, and also whether performance of tasks is improved by offering formal performance-related rewards (PRR) rather than fixed rewards (FR); see e.g. Tuttle and Burton (1999), Waller and Chow (1985). For the decision-facilitating role, empirical tests of theoretical predictions concerning the use of information by a single individual have shown that individuals do not generally choose efficiently between information systems and that they also do not process information in ways consistent with conventional theoretical models (Bonner, 1999; Waller, 1995). For the decision-influencing role for information in multi-person organisational environments, researchers have found that, when compared to FR contracts, PRR contracts linking rewards to monitoring variables do tend to motivate individuals to improve their performance in a range of tasks (Sprinkle, 2003).

A common environment for employees with decision-making authority is one in which the two roles for information are combined. Employees have access to information that could enhance their performance of a task *and* their task performance is monitored. Thus an interesting question relating to the use of formal performance-related rewards as well as to the provision of accounting information, in these more complex yet widespread environments, is whether formal performance-related rewards enhance the use decision-makers make of valuable decision-related information and therefore improve task performance. To put the issue another way, the two functions of information may not be independent: if an employee could improve decision-making by accessing and utilising information, but needs to be motivated to do so by the formal linking of subsequent extrinsic rewards to some measure of performance, then it is reasonable to suppose an employee with PRR would make better use of decision-facilitating information, and on average also perform better as a result, than an employee with FR.

There are relatively few experimental studies, relating to this *dual* informational role, published in the accounting and social psychology literatures, and results so far have been rather mixed. For subjects working in an investment analysis context, Tuttle & Burton (1999) found that those with PRR used more information for the task than those with FR and they performed better. Drake et al (1999) found that group-based rather than tournament-based incentives improved group engagement with an activity-based costing system. Sprinkle (2000) found that feedback on past performance level was more likely to be used to improve future performance if subjects were rewarded with PRR rather than FR. On the other hand, Arkes et al (1986) found that subjects facing a probabilistic task were less likely to use a helpful decision aid when given PRR; in Ashton (1990), subjects engaged in a bond-rating task, and with access to a helpful decision aid, performed worse with PRR than with no monetary incentives; Hogarth et al (1991) found, if penalties for mistakes were relatively large, subjects with PRR performed less well on a repetitive task where feedback was available; and finally, in Ravenscroft & Haka (1996), information sharing, productivity gains and variance reduction were not improved by the provision of competitive PRR.

A consensus view on the meaning of these disparate results is problematic, partly because of differences between studies in experimental environments selected for examination, and partly because of alleged deficiencies in some of the experimental designs employed. On this latter point, for the four studies finding no improvement in task performance, Sprinkle (2000, 2003) argued that their use of tournaments and similar incentive structures as rewards would lead to experimental subjects adopting high-risk strategies not necessarily consistent with maximising expected performance. Furthermore, he questioned the quality of the feedback information provided to subjects in three of the four studies, implying that subjects who chose to ignore it when making decisions could not be described as unmotivated. The experiment reported in Sprinkle (2000) was explicitly designed to improve on these alleged deficiencies and, as noted above, it found that PRR tended to enhance task performance.

The objective of the present work was to replicate, extend and study the robustness of the findings reported in Sprinkle (2000). The value of replication comes from discovering whether experimental phenomena are robust to large or small changes in experimental design, or whether they are more or less unique to a specific design; see Guala (2005).[1] In the present context, it is of interest to investigate whether the main findings of Sprinkle (2000) are robust to variations in detail of

[1] Research in the 'social' as opposed to the 'natural' sciences has a tendency to focus on novelty, a tendency to undervalue the importance of replication. For instance, Smith (1994) and Rubinstein (2001) have argued persuasively that replications are not valued sufficiently highly in economics-based research. The tendency may be understandable in some areas, but in experimental work it is our view that the standards of the natural sciences can and should apply – and that there is significant value in replication *per se*. That said, and as explained in what follows, the present work is not confined to replication *per se*.

implementation in the laboratory, and with a rather different cohort of participants; Sprinkle's work was based on a sample from undergraduate US university programmes, whereas in our study we drew on both undergraduate and postgraduate students, with an international but non-US educational background, primarily from Europe and Asia. If the phenomena found in Sprinkle (2000) persist, then increased confidence can be gained that PRR 'works'. Computerisation of the experimental environment also facilitated dealing with a significantly increased sample size relative to Sprinkle (2000).

Our research question remained broadly the same as in Sprinkle (2000):

> Relative to fixed incentives, does provision of performance-related incentives increase the use made of feedback information on past performance and enhance task performance?

However, there were significant differences between the present work and Sprinkle (2000) in how research hypotheses were formulated, in how the experiment was implemented, and in how the data were analysed. Two of these differences arose as a response to a pilot study undertaken for the present work, where it was found that PRR did increase subject information usage and enhance task performance compared to FR but, in contrast to Sprinkle (2000), the effects were weak and insignificant.[2] Indeed, FR participants did nearly as well as PRR participants. Two issues potentially contributing to the weakness of these results emerged from the pilot study: saliency of incentives and subject comprehension. Given their importance to our experimental design, these issues are discussed now, with discussion of other minor differences between the present paper and Sprinkle (2000) deferred to Section 2.

With regard to saliency, Davis & Holt (1993, p 24) maintain 'it is critical that participants receive salient rewards that correspond to the incentives assumed in the relevant theory or application.' Many experimental applications assume participants have no intrinsic incentives, either positive or negative, regarding performance of the experimental task; see Plott (1982). If that assumption is valid, then it is straightforward, at least in principle, to align incentives with the theory being tested through provision of arbitrarily small money payments to experimental subjects. But the assumption of no intrinsic preferences is rarely satisfied. Onerous tasks may involve intrinsic costs to subjects or, conversely, a desire to achieve success in a task may motivate subjects to try hard even when

[2] Detailed results of the pilot study are reported in Dobbs & Miller (2006).

there are no money payments for good performance; see Rieken (1962). There is some experimental evidence that the conformity of experimental behaviour with theoretical models is affected by saliency of reward: see for example Siegel & Goldstein (1959), Hirshleifer & Riley (1992) and Prasnikar (2002). Thus, Davis & Holt (1993, p 24) recommend *salient* rewards, which means:

> '*subjects perceive the relationship between decisions made and payoff outcomes, and… the induced rewards are high enough to matter in the sense that they dominate subjective costs of making decisions…*'[3]

In the pilot study for the present work, it was clear FR participants made significant use of feedback information in order to improve their task performance, despite the fact that their explicit financial rewards gave them a small disincentive to behave in this manner. This observation suggests that participants had personal preferences in favour of achieving a better performance in the experimental task, and that the level of financial disincentive was simply *not salient enough* to dominate these intrinsic personal preferences or to induce significant measurable differences between PRR and FR participants.[4] In other words, it seems there was little difference on average between the behaviour of the two groups because both were similarly intrinsically disposed towards succeeding in the task set for them. When subjects have unknown levels of intrinsic preferences, the calibration of explicit rewards necessary to align incentives with the relevant theory is essentially an empirical issue. In the present work, instead of attempting to discover critical levels of cost disincentive for inducing the FR group to eschew opportunities to access information, we chose to examine the empirical effects of saliency on subject behaviour in a more direct and general manner.[5] Thus we varied both the benefits and costs of accessing information across nineteen separate treatments, rather than the two treatments with fixed costs as used in Sprinkle (2000) and our pilot study. Our working research hypotheses became:

> H_1: Other things equal, as the level of performance-related payment increases, subjects access information on past performance more and their profit performance is enhanced.

---

[3] Equally, the words 'subjective benefits' could be used in place of 'subjective costs', depending on the experimental context.

[4] In the pilot study the magnitudes of money incentives were selected to approximately match those used in Sprinkle (2000). The financial disincentive to accessing information was relatively small at £0.008 for each information request.

[5] Even if the pilot study had confirmed the findings of Sprinkle (2000) with relatively small money incentives, it has been argued that an examination of saliency would still be warranted in order to investigate whether observed behaviour survives when the 'incentives for thinking things through carefully' are increased; see Binmore (1994, pp.184-5).

H_2: Other things equal, as the cost of accessing information increases, subjects access information on past performance less and their profit performance is worsened.

The nineteen levels of incentives we employed included the two treatments tested in our pilot study. Hence, the original test in Sprinkle (2000) and in our pilot study for differences between PRR and FR subjects was retained here as a nested hypothesis of the more general model. However, unlike Sprinkle (2000) and Dobbs & Miller (2006) where the cost of information was fixed, information is also provided on the effects of varying the cost of accessing information.[6]

With regard to subject comprehension, the post-experiment questionnaire used in the pilot study revealed some variation across participants in their apparent comprehension of aspects of the experimental environment. Observations on the behaviour of subjects who do not understand the presented laboratory environment cannot be unambiguously attributed to that environment. For this reason, laboratory studies finding apparently imperfect comprehension of the experimental environment have tended to exclude such subjects from subsequent data analysis, though they also often provide a footnote if the results are not sensitive to inclusion of these subjects; see for example, Davis et al (2006). For complex tasks and environments, however, it is not always obvious how to judge the significance of errors in responding to such questionnaires. Depending on the complexity of the questions posed and the incentives faced (typically none, when responding to such questionnaires), whilst errors *may* reflect imperfect understanding, errors *may* also result from calculation or transcription errors as well as errors in conceptual understanding. Indeed, the more extensive and onerous the questionnaire, the higher the number of respondents that can be expected to make some errors. Given the fact that the post experiment questionnaire is inevitably to an extent an imperfect 'discriminator' of levels of understanding, the natural response is to explicitly incorporate the questionnaire data within the statistical model. The idea is that if a subject does not understand some aspect of the experimental environment, then they will not be able to take it into proper account within the experiment – and so it should be possible to formally test whether this is so or not. In effect, we suggest that the 'data should decide' whether groups with apparent different levels of understanding (as revealed by post experiment questionnaire) can be pooled. That is, instead of assuming for the purposes of analysis that data from subjects is either of no value or full

[6] The response variable 'time spent on the task' in Sprinkle (2000) was here replaced by 'number of information requests'. We focused upon the latter variable because it better reflects demand for information. In Sprinkle (2000) the two variables were related by design. Removing this relation and analysing the two variables separately showed, for our sample, that time spent on the task was unaffected by treatment; this point is discussed in more detail in our ICAS report, Dobbs & Miller (2008).

value, to be excluded or pooled respectively, we allowed the data to determine statistically whether pooling is appropriate or not.

Our approach is thus a novel but also a methodologically consistent way of making use of post-experiment questionnaire data. Firstly, it should be noted that full comprehension of the experimental environment by subjects is only a sufficient condition for a valid experiment – it is not a necessary condition. As a consequence, it is not necessarily theoretically correct to assume that data from subjects with imperfect comprehension is of *no* value in subsequent analysis. At the same time, it is also not necessarily theoretically correct to assume such data is of *equal* analytical value to data from subjects with full comprehension. Secondly, it is important to recognise that tests of understanding are inevitably imperfect discriminators of comprehension[7] and that exclusion could involve substantial wastage of data. Indeed, it can be expected that the number of subjects a Davis et al procedure leaves standing depends on the complexity and extensiveness of the questionnaire subjects are required to complete. Given these considerations, our approach improves on the 'exclusion or pool' dichotomy of Davis et al (2005) by means of explicit statistical modelling. The value of our approach becomes clear in Section 3 below, where it yields some interesting insights that would otherwise have been obscured.[8] For completeness however, and for comparison purposes, we also present results with full exclusion, based on the sub-sample of 42 subjects who answered the questionnaire fully correctly.

A final difference between research hypotheses employed in the present paper and in Sprinkle (2000) was the inclusion here of demographic variables.[9] The experimental task required of each subject was the solution of a series of problems, or *de novo* puzzles. There is now a substantial body of evidence indicating that skill in accomplishing such tasks depends upon gender, age and grade point average (GPA); see for example Hambur et al (2002), Lundeberg et al (1994), and Research and Library Services, Northern Ireland Assembly (2001).[10] In particular, for problem-solving tasks males do better than females and older subjects do less well than younger subjects. Dermer (1973) has argued for inclusion of demographic variables in accounting experiments, especially when the

[7] For example, even with full understanding, computational and other types of mistakes can occur which result in a 'wrong' answer.
[8] That the behaviour of subjects with less than full comprehension of a laboratory task can suggest useful questions for further research is also demonstrated by studies of the preference reversal phenomenon, which provides evidence on issues relating to human cognition. See, for example, Starmer & Sugden (1991).
[9] Sprinkle (2000) does not report demographic information or analysis.
[10] GPA is generally correlated with other measures of achievement or skill, including problem solving; see Hambur et al (2002) and Kuncel et al (2004).

influence of demographic factors has not been removed by randomisation.[11] Hence, in order to statistically control for such effects on task performance, personal characteristics of subjects are included in the analyses. For the present study, the published literature suggests three further research hypotheses:

> H_3: Other things equal, the profit performance of males is better than females.

> H_4: Other things equal, profit performance is negatively correlated with age.

> H_5: Other things equal, profit performance is positively correlated with GPA.

Data were also collected on nationality and graduate status. For nationality, subjects were coded into four categories: British, Non-British European, Asian and Other. Anecdotally, it is often stated that Asian educational processes emphasise the memorising and reproducing of facts rather more than problem solving and creative thinking. [12] Hence, an additional research hypothesis considered was:

> H_6: Other things equal, the profit performance for Asian subjects is worse than for other subjects.

A hypothesis for the graduate status variable, undergraduate or postgraduate, is rather more difficult to formulate for this task, because graduate status will be positively correlated with both age and GPA, and the effects of these two variables are hypothesised to work against each other. On balance, however, the latter effect may be expected to predominate, hence:

> H_7: Other things equal, the profit performance for postgraduate subjects is better than for undergraduate subjects.

---

[11] Deux (1979) and Ciancanelli et al (1990) provide good examples of the use of demographic analysis outside the laboratory, in studies of gender experience in the workplace and in the accounting profession respectively.

[12] A referee has pointed out that there exist some difficulties of interpretation with regard to nationality. When eliciting information from subjects, we chose 'nationality' rather than a cultural label, in order to feel confident that, for example, British nationals with an Asian family background, who have come up through the British educational system, would predominantly label themselves as 'British'. Hence subjects in the 'Asian' category are of overseas origin, in fact overwhelmingly Chinese, with a somewhat different educational experience compared with that of British subjects.

For ease of comparison, the research hypotheses of the present paper and Sprinkle (2000) are summarised in Table 1 below. The table highlights many of the areas in which the present paper contributes to the literature. Firstly, we attempt to replicate the finding in Sprinkle (2000) that PRR increases the use made of valuable but costly information and consequently enhances task performance relative to FR. This finding is related to the theory of incentives from agency theory and has direct application to the design of performance evaluation and reward systems based on accounting data. Our results offer additional support that PRR works as predicted. Secondly, we extend the two-treatment experimental design used in Sprinkle (2000), varying the incentive levels more widely across nineteen treatments in order to examine the impact of salience on the demand for information and on task performance. Our results show that the findings in Sprinkle (2000) are robust to the level of PRR, which seems to exert a continuous impact on the response variables. Thus the results provide evidence that the effect of PRR is not restricted to a special case and that potentially the theory of incentives has wide applicability to observed behaviour. Our manipulation of information cost, not considered in Sprinkle (2000), produced results that indicate subjects access information more often when it is less costly but that task performance is not significantly affected.

Table 1: Comparison of Research Hypotheses in Sprinkle (2000) and Present Study

Sprinkle (2000)	Present Study
Average time spent on task will be greater for subjects receiving PRR than for subjects receiving FR (2 treatments)	Number of information requests will be increasing (decreasing) in level of PRR (cost of information) (19 treatments)
Average profit performance will be greater for subjects receiving PRR than for subjects receiving FR[13] (2 treatments)	Average profit performance will be increasing (decreasing) in level of PRR (costs of information) (19 treatments)
	Average profit performance will be greater for males
	Average profit performance will be decreasing in age
No corresponding tests reported	Average profit performance will be increasing in grade point average (GPA)
	Average profit performance will be lower for Asian than for other subjects
	Average profit performance will be greater for postgraduate than for undergraduate subjects

[13] Sprinkle (2000) also examined a third research hypothesis: that the average positive difference in profit performance for subjects receiving PRR compared to FR would increase with repetition. This hypothesis was originally included in order to explain why previous single-trial comparisons of profit performance had failed to find a positive average difference for PRR subjects. In the present work, we focus on performance post learning effects, therefore on statistics from the latter half of the experiment, after subjects had obtained experience of the task and incentives.

Thirdly, the demographics of our sample differed from those in Sprinkle (2000). The confirmation that PRR works across a wider range of subject types suggests greater external validity for the results. Our evidence suggests that gender and cultural variables affect both behaviour and task performance for this task. Fourthly, a novel feature of the present study is our inclusion in the statistical model of subjects with apparently imperfect comprehension of the experimental task and incentive structure. Dummy variables are used in combination with a standard general-to-specific testing down method to estimate the effect of including such subjects in the analysis. Some potentially interesting results emerge: all subjects respond in the same degree to variations in both cost and PRR level, by varying their demand for information in the predicted directions, but this translates into the predicted variation in profit performance *only* for subjects with apparently perfect comprehension of the experimental environment. This distinction would not have been apparent if the affected data had been fully excluded from the analysis, for the generality in response for all subjects in information usage would have been missed.

The rest of the paper is organised as follows. Section 2 details the experimental environment in which data were generated and organised. Section 3 discusses the statistical analysis and presents the results. Section 4 concludes.

2. The Experimental Environment

The decision-making task faced by participants was based on Table 2.

Table 2: The Relation between Participant Decisions and 'Profit' Outcomes

		Column Choices																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	5	5	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	5	5	10	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	5	5	10	20	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	5	5	10	20	20	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	5	5	10	20	20	30	30	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	5	5	10	20	20	30	30	30	0	0	0	0	0	0	0	0	0	0	0	0
	9	5	5	10	20	20	30	30	30	45	0	0	0	0	0	0	0	0	0	0	0
Rows	10	5	5	10	20	20	30	30	30	45	45	0	0	0	0	0	0	0	0	0	0
	11	5	5	10	20	20	30	30	30	45	45	60	0	0	0	0	0	0	0	0	0
	12	5	5	10	20	20	30	30	30	45	45	60	60	0	0	0	0	0	0	0	0
	13	5	5	10	20	20	30	30	30	45	45	60	60	60	0	0	0	0	0	0	0
	14	5	5	10	20	20	30	30	30	45	45	60	60	60	80	0	0	0	0	0	0
	15	5	5	10	20	20	30	30	30	45	45	60	60	60	80	80	0	0	0	0	0
	16	5	5	10	20	20	30	30	30	45	45	60	60	60	80	80	95	0	0	0	0
	17	5	5	10	20	20	30	30	30	45	45	60	60	60	80	80	95	95	0	0	0
	18	5	5	10	20	20	30	30	30	45	45	60	60	60	80	80	95	95	95	0	0
	19	5	5	10	20	20	30	30	30	45	45	60	60	60	80	80	95	95	95	100	0
	20	5	5	10	20	20	30	30	30	45	45	60	60	60	80	80	95	95	95	100	100

This profit table is identical to that used in Sprinkle (2000). Both columns and rows in Table 2 are labelled 1,…,20. The remaining numbers are regarded as profit outcomes for the firm associated with particular combinations of row and column. For example, the profit associated with Column 14 and Row 19 is found, at the intersection of that column and row, to be 80. Each participant was given a copy of Table 2 and informed of the reward they would receive per unit of profit earned for the firm, as depicted in Table 2, and also the cost per information request they made. They also knew the experiment was broken into 12 trials each of which comprised 5 periods.

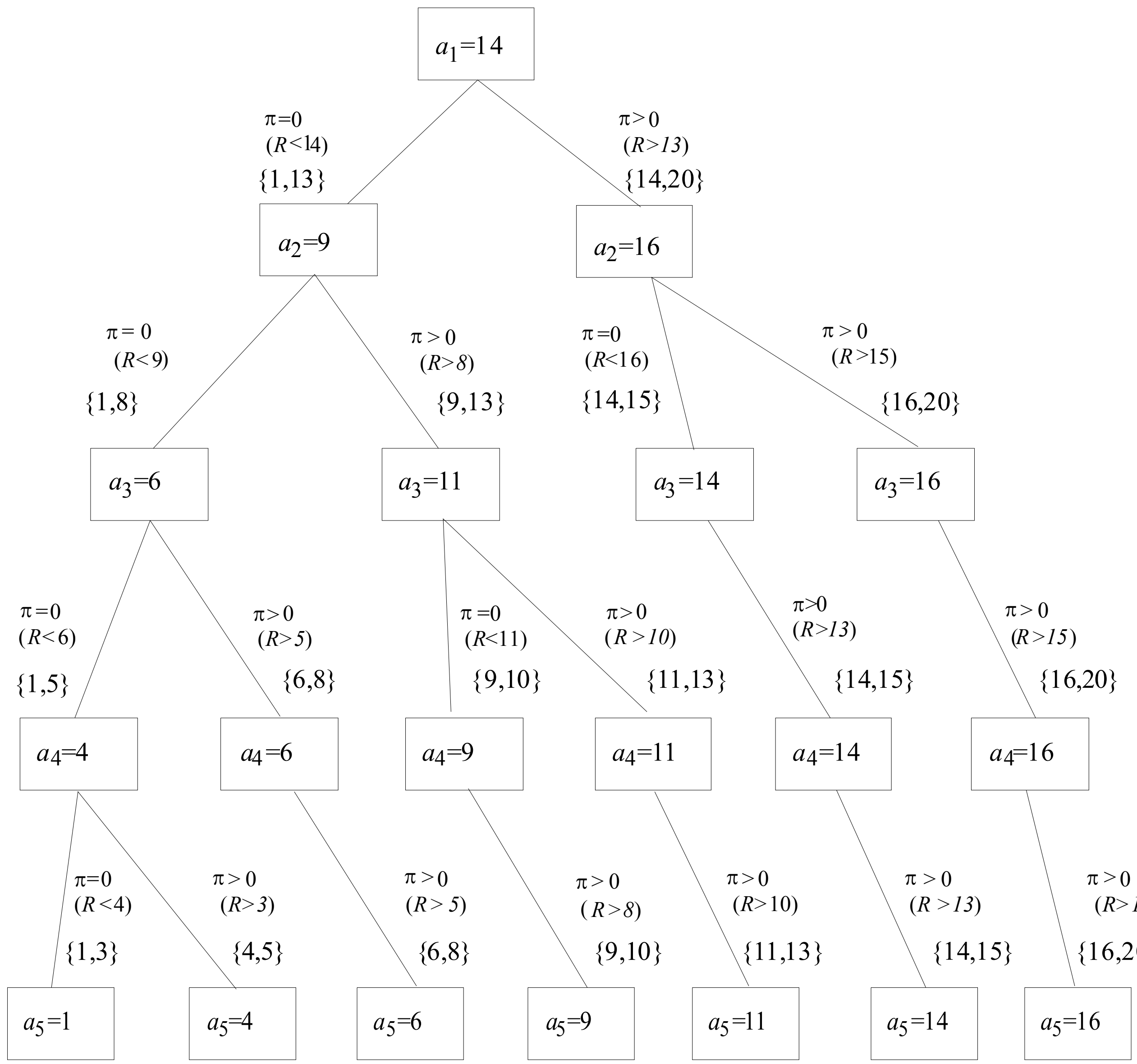
At the beginning of each trial, a row in the above profit table was randomly selected by computer. Participants were informed that the computer had been programmed so that each row was equally likely to be selected, and also that the row for one participant was selected independently of the row selected for any other person participating in the experiment, and independently of all decisions they themselves took during the experiment. The selected row was not revealed to the participant, who merely knew that there was an equal chance of any given row being selected. However, the individual was informed that the row selected would remain the same for the next 5 periods of the trial. In each period, the participant was required to make a choice of column. Following this, they were asked if they wished to make an 'information request' in order to find out what profit had been earned by their choice. If they made a request, they were informed of the profit earned for the firm, as well as the cost they had incurred by making an information request. This ended the period, and the individual moved on to the next period. If no information request was made, the individual moved directly to the next period without being given any information. At the end of 5 periods, the trial was complete, and the individual was given a break down of the incentive pay earned and information costs incurred in the trial. The process then repeated; that is, in the next trial, a new row was randomly selected by the computer and then fixed for the duration of the trial, the individual starting the process of choosing columns and making information requests all over again. The decision problem faced in each of the 12 trials was thus the same, except for the randomly-selected row. The rubric for the experiment asked all individuals to try to maximise the firm's profits. Their individual money incentives, which varied across individuals, were revealed privately to reduce the possibility that subjects would know the research hypotheses being tested; see Rosenthal (1963).

Let Π stand for the profit outcome, *C* for column choice and *R* for the row. Notice that each column of the profit function in Table 2, except Column 1, has two possible profit outcomes; either $\Pi = 0$ or $\Pi > 0$. The design of the profit function is such that, for any given column, the $\Pi = 0$ and $\Pi > 0$ profit outcomes partition that column into two contiguous non-overlapping sets of rows. For example, for Column 13, Rows 1 to 12 generate zero profit and Rows 13 to 20 generate a positive profit of 60. Thus knowledge of the profit outcome provides additional information about the undisclosed row which had been fixed by the computer at the start of the trial. Suppose in the first period of the experiment a participant did choose Column 13. If, following this, the participant does not choose to learn the resulting profit outcome, then the uncertainty about the row selected will be the same as it was before the choice of column; that is the row could be any of Rows 1 to 20; $R \in [1, 20]$. However, if the participant does choose to learn the resulting profit outcome, and it turns out to be zero, then the participant for certain can rule out all of Rows 13 to 20 as possibilities; that is $R \in [1,12]$. By contrast, if the request for information revealed that profit was 60, then the participant for certain can rule out all of Rows 1 to 12 as possibilities, so $R \in [13, 20]$. Hence, for this choice of Column 13, the participant can for certain learn, after this first period, something they did not know before. The same point holds for *any* choice of column except Column 1, since Column 1 offers a constant profit of 5 whatever the selected row.

Pursuing the above example, of a choice of Column 13 in Period 1, a little further, suppose the information request revealed profit to be zero. Then the participant knows for certain that the same column choice or indeed any higher-numbered column will produce exactly the same zero profit for the remaining four periods of the trial. The participant wishing to maximise profits would thus rationally choose a lower-numbered column in the next and subsequent periods of the trial. To put it another way, it is clearly irrational to choose a column $C \in [13, 20]$ in any subsequent period in the trial. By contrast, if the information request revealed a profit of 60, this means it would be irrational to choose $C \in [1,12]$ in any subsequent period of the trial, and this knowledge might encourage the participant to increase the column number chosen in the next period in the search for even higher profits. Note that, in this case, one can always 'retreat' back to Column 13, should a higher-numbered column choice subsequently produce a zero profit.

Formally, maximising expected profit constitutes a dynamic programming problem of some complexity, one which is quite impossible to solve analytically in the time available to participants. A flowchart of optimal decision paths is shown in Figure 1 below. Notwithstanding the complexity of the problem, a 'rational' individual should make choices which give a positive probability of positive profit; that is, they should not choose columns that previous information requests have indicated will guarantee zero profit. Also, they can be expected to operate some kind of 'rule of thumb' in which the initial choice is a column somewhat above Column 11 and, following the initial information request, tend to edge up the column choice, if revealed profit was positive, or edge down, if revealed profit was zero, a process that can be repeated period by period throughout the trial. It is worth noting that there is never any decision-facilitating value to learning fifth-period profit, since at that point the trial is complete and the participant either finishes the experiment or moves onto a new trial, with a new independently-generated row number.

Figure 1: Optimal Decision Flowchart



Key to Figure 1:

a_i = optimal choice of column in period *i*

$\pi = 0$ or $\pi > 0$: profit is then revealed as zero or positive.

$R < Num$ or $R > Num$: Inference then possible concerning the number of the row selected by the computer.

$\{x,y\}$: Given column choices and information requests, it is then known that the computer's choice of row lies in this range; that is $x \leq R \leq y$.

Optimal play involves choosing column 14 at period 1 ($a_1 = 14$) and then asking for information and if profit is positive, moving up to $a_2 = 16$, or if profit is zero, to $a_2 = 9$ and so on.

To explore the role of saliency on the motivational impact of PRR, participants were provided with a much wider range of incentives and disincentives than in Sprinkle (2000). In particular, in the present paper the money reward to participants per unit of profit earned was varied between £0.00 and £0.02 in steps of £0.005, and the money cost to participants per information request was varied first from £0.01 to £0.05 then up to £0.20 in steps of £0.05. To minimise the possibility participants might end up with negative money rewards from the experiment, the upper bound for varying the cost per information request was reduced when money reward per unit of profit earned was small. FR participants, with rewards unrelated to profit performance, were given a fixed money reward per column choice made, but information costs were then systematically varied over the full range. Volunteers were randomly assigned to incentive conditions, except for the set of FR assignments, which were assigned only in the first session. With these ranges of rewards and costs, average participant payouts from the experiment were £19.50, compared to an average of £9.85 in our earlier pilot study. Table 3 shows the sample frequencies of each assigned combination of reward per unit of profit earned and cost per information request.

Table 3: Number of participants with given reward/information request incentive conditions

		Reward per Profit Point					Totals
		0	0.005	0.01	0.015	0.02	
Cost per Information Request	0.01	4	5	6	4	4	23
	0.05	7	6	6	4	4	27
	0.1	7	-	4	5	3	19
	0.15	4	-	-	5	5	14
	0.2	6	-	-	-	5	11
Totals		28	11	16	18	21	94

Four experimental sessions were conducted at Newcastle University during the spring and early summer of 2006. For each session, an instructor attended one of the target cohort's lectures, handed round copies of a 'Personal Details Form', and asked for volunteers to participate in the experiment. The 'Personal Details Form' provided details on sex, nationality, age, degree course and contact email for those wishing to volunteer, as well as a consent declaration allowing the researchers to use data on end-of-year grades in subsequent statistical analysis. A total of 94 participants provided potentially useable data. Volunteer participants were drawn from various degree programmes involving business, economics and accounting, with a mix of undergraduates and postgraduates, home and overseas, male and female students. All those available for the session dates and times were accepted for participation. Table 4 presents sample descriptive statistics.

Table 4: Sample Descriptive Statistics

% Male	56
% Postgraduate	35
Average Grade (%)	61
Average Age (Years)	22
% British	62
% Non-British European	7
% Asian	22
% Other Nationality	7

At the start of each experimental session, participants first received instructions in a classroom, and then moved to the computer cluster, where each member of the group sat at an individual terminal to undertake the experiment. In the classroom, an instructor read out, word for word from a prepared document, instructions implementing the experimental environment. Participants could follow the instructions from their own individual hardcopies, which included Table 2, provided just before the session began but after initial rules of communication were explained. The instructions for PRR participants appear here as Appendix 1; the instructions for FR participants were similar. The instructions were drafted and the sessions arranged so that no participant knew the variation in treatment faced by any other participant, and no participant had information about hypotheses under examination. The instructions in Appendix 1 carried only a neutral heading, without indication of treatment, and the individual treatments were disclosed only when participants were physically separated from each other at computer terminals, about to begin the experiment. A computer program was specially written to create the experimental environment. Further information concerning the Personal Details Form, the instructions for FR subjects, the computer program, and a selection of screen-shots, can be found in Dobbs & Miller (2008).

The post-experiment questionnaire, included here as Appendix 2, formed the basis for investigating whether profit performance was influenced by variation in a subject's level of understanding of the experimental environment. Since the questionnaire was administered following completion of the experiment, it can be considered to have had no effect on decision-making *during* the experiment. Some studies attempt to avoid the problem by means of compulsory ex ante testing of individuals, followed as required by reinforcement of instructions; see, for example Fisher et al (2002) and Fisher et al (2003). However, details of the test are typically not included in published documentation accompanying the results, making it difficult to assess the extent to which the procedure is effective, either in particular cases or more generally. Other studies have relied on ex ante opportunities for individuals to pose their own questions to an instructor; see Plott (1982). The conventional wisdom on this is that all questions about the experimental environment should be met with repetitions of the relevant portions of the written instructions, thus reducing the possibility of an instructor being drawn into unwittingly revealing clues about expected or desired responses from participants, particularly using unscripted remarks that may then be difficult to document and therefore replicate; see Rosenthal (1963). The ability or confidence of participants to frame suitable questions, either in private instruction or publicly during group instruction, is open to doubt however, as is the general effectiveness of an instructor who responds to questions only by repeating the original instructions. The post-experiment questionnaire has recently been employed by Davis et al (2006), who subsequently reported results that excluded 6% of their sample data; though they also reported their findings were not affected if results are based on their full sample. Typically, post experiment questionnaire data has not been explicitly used as an integral part of the statistical modelling procedure. Rather it has been used, as in Davis et al (2006), to rationalise the exclusion of participants with imperfect comprehension. As explained in section 1 above, in our view this is an unduly draconian approach. In the present work, the post experiment questionnaire is therefore used to identify different levels of *apparent* comprehension of the experimental environment. A general-to-specific methodology is then used to identify the extent to which these distinct sub samples can be pooled.

The questionnaire consisted of five questions. Questions 1 and 2 tested understanding of aspects of the profit table; respectively, whether a participant could correctly ascertain profit from a given row and column, and could infer the range of unobserved rows consistent with both a given column and profit report. A total of 70 participants correctly answered both Questions 1 and 2. This group was assigned a code of ‘G1’, denoting Group 1, whereas the full sample was coded ‘G0’. Question 5 tested understanding of the individual cost to a participant of accessing feedback information. A total of 62 participants answered all of questions 1, 2 and 5 correctly. This group, a proper subset of group ‘G1’, was coded ‘G2’. Finally, Questions 3 and 4 dealt with individual participant rewards and their relation, if any, to earned profit. Only 42 participants correctly answered all of questions 1 through 5. This group, a proper subset of group ‘G2’, was coded ‘G3’. Hence, prior to undertaking statistical analysis, the participants were ‘filtered’ into 4 nested groups, ‘G0’, ‘G1’, ‘G2’ and ‘G3’.

The idea behind the nested classification was to provide a simple and parsimonious ordering of the apparent level of 'understanding', as this might impact on ability to earn profit. 'Raising the bar' for inclusion in each successive higher-numbered grouping allows ordinal interpretations of profit-earning capability for the resulting four groups, and use of a general statistical model for estimating the particular response relating to each level of understanding as a nested special case of the more general model.[14]

It is an empirical question as to whether sub-groups can be pooled or not, but it is worth emphasising that the questionnaire responses are only *indicative* of the level of understanding of participants. The post experiment questionnaire-based classification of participant understanding is not claimed to be perfect. Assessment was based upon a small number of questions as well as subjective appraisal of the significance of apparent error. Moreover, completion of the questionnaire was not linked to any money consequences for subjects, so the extent to which it was taken seriously, apart from through a desire to cooperate and to display achievement, might be doubted. Whilst a correct answer probably does reveal good understanding, an incorrect answer to a question on the post experiment questionnaire may or may not reveal a lack of understanding – as in any exam, a participant may have full understanding and yet make a mistake.

As well as attempting to estimate the effects of apparent differences in participant understanding of the task and incentives, we also sought to reduce the complexity of the environment from the one used in Sprinkle (2000) and our pilot study, cutting out any unnecessary distractions. Firstly, in Sprinkle (2000) a lottery procedure designed to 'induce' in all participants a neutral attitude to risk was employed; see Berg, Daley, Dickhaut & O'Brien (1986). Selten et al (1999) have argued that the lottery procedure adds complexity and tends to reduce subject comprehension of the experimental environment. Given the relatively small subject rewards involved, and the fact that rewards are always positive, there are good grounds for assuming that behaviour is likely to be approximately risk neutral; the logical case for this is explained clearly in Arrow (1970) and Rabin (2000). Thus in the experiments reported in this paper, the lottery procedure is no longer included as part of the experimental design, and rewards are earned directly in money. Secondly, in Sprinkle (2000), there was a very small time-related penalty: taking longer to complete the whole experiment reduced by a small amount the overall reward a subject earned. We dispense with this penalty in the

[14] It is possible to conceive of many other classification schemes; for instance, with five questions it would have been possible to divide participants into 2^5=32 separate non-nested classifications. However, such a proliferation of groups would have resulted in uneconomically large estimation demands on the available data.

current experiments, because our focus is not on *time-pressured* decision-making.[15] The differences in experimental environments in Sprinkle (2000) and the present paper are summarised in Table 5 below. Note the more generalised analysis possible in the present study, through the inclusion of a significantly greater number of reward/cost treatments, see also Table 3, and the wider coverage of rows assigned to participants, see also Table 2.

Table 5: Comparison of Experimental Environments in Sprinkle (2000) and Present Study

Sprinkle (2000)	Present Study
Identical decision-making task with stationary stochastic profit function	
2 levels of reward/cost parameters	19 levels of reward/cost parameters
Rewards/costs denominated in probability points for binary lotteries	Rewards/costs denominated in money
9 computer-generated rows examined	20 computer-generated rows examined
Small penalty assigned to time spent on task	No time-related penalty

3. Statistical Analysis

This section details the development of a statistical model for testing research hypotheses H_1 through H_7, as applied to the response variable, trial profit.[16] Given the structure of the profit table, Table 2, it is clear that total profit over the 5 periods of a single trial is wholly determined by the single row randomly selected by the computer, indexed by R_{it} for individual *i* in trial *t*, and the set of five column decisions taken by individual *i*. In turn, research hypotheses H_1 and H_2 postulate that a participant's column decisions will be influenced by the formal incentives supplied; respectively, the real money reward to the participant per unit of profit earned, denoted r_i^{π}, and the real money cost to the participant per request for feedback profit information, denoted c_i^{IR}. An individual participant, faced with these incentives, will adopt some strategy for selecting columns. For fixed r^{π} and c^{IR}, research hypotheses H_3 through H_7 predict that this strategy may differ across individuals, according to various personal characteristics, such as age, sex, graduate status,

[15] Sprinkle (2000) noted that his results were not sensitive to the inclusion or exclusion of time-related penalties.
[16] The development of the model for the response variable 'number of information requests' is exactly analogous and is therefore omitted.

nationality and intellectual ability, the latter proxied here by the most recent grade point average, *gpa*. A participant's strategy may also change during the course of the experiment as experience, indexed by trial number, is gained with the decision-making task. Adding a random term, ε, with average value zero, to account for variables not specifically articulated in the model, it is then possible to specify a reduced-form model of the determinants of earned trial profit π, which depends on nine measurable variables and a random term, as in equation (1). Here, *mf*, *upg* and *nat* are indicator variables for sex, undergraduate or postgraduate status, and nationality, respectively; while *f* denotes the, as yet unspecified, form of the functional relationship:

$$\pi_{it} = f(R_{it}, r_i^{\pi}, c_i^{IR}, age_i, mf_i, upg_i, nat_i, gpa_i, t) + \varepsilon_{it} \qquad (1)$$

for individuals $i = 1,..,94$ and trials $t = 1,..,12$.[17] The strategy employed in this paper was to specify a particular functional form for equation (1) and then estimate it using OLS regression.

Many, though by no means all, of the standard parametric hypothesis tests that could be performed on specific versions of equation (1) rely on an assumption that u is a linearly independent, Normally-distributed random variable, with constant variance. For certain specifications of equation (1), the validity of this assumption is open to doubt. For example, in equation (1), π measures the profit made by an individual in a single trial. Since the profit table row is held constant for the duration of each trial, this means that to appropriately apply standard parametric tests based on Normal distribution theory, the distribution of trial profit must be Normal for every individual profit table row, for arbitrary values of the other eight determinants. Table 6 lists feasible values of π, for each profit table row.

[17] In Sprinkle [2000], c_i^{IR} was fixed at a constant level and therefore could not influence variation in actual trial profit, while r_i^{π} was a dichotomous variable. Because there were only 2 treatments to be compared in Sprinkle (2000), it was feasible to divide up the sample by profit table row or trial, and to then conduct pairwise difference-of-means tests. In the present study, however, with 19 distinct incentive pairs for r^{π}, c^{IR}, this procedure was not practical.

Table 6: Feasible values for trial profit

Profit Table Rows	Feasible Values for π
1-2	0, 5,… 25
3	0, 5,… 50
4-5	0, 5,… 90, 100
6-8	0, 5,… 130, 140, 150
9-10	0, 5,… 190, 200, 210, 225
11-13	0, 5,… 250, 260, 270, 285, 300
14-15	0, 5,… 330, 340, 350, 365, 380, 400
16-18	0, 5,… 390, 400, 410, 425, 440, 460, 475
19-20	0, 5,… 410, 420, 430, 445, 460, 480, 495, 500

Clearly, π is a discrete variable, with a row-dependent upper bound and a lower bound of zero. In contrast, a Normally-distributed variable is continuous and unbounded. The problem is most severe for low-numbered rows, where there is likely to be a high proportion of observations at the lower bound, potentially affecting the tests employed in Sprinkle (2000). Although some writers claim standard tests are robust to departures from Normality, see for instance Clinch & Keselman (1982) and Tan (1982), others have opposing views; see Bradley (1978), Glass et al (1972), Wilcox (1995) and Wilcox (1998). A second issue to address in equation (1) is its panel data, or repeated measures, aspect; if π is trial profit, then there are twelve profit observations for each participant. Ignoring this aspect may produce linear dependence between profit observations; see Maddala (2001). Fortunately our primary focus in what follows is solely on whether the incentive condition faced by an individual influences behaviour and performance. This can be addressed by aggregating the data across trials for each individual. Specifically, the profit response variable can be re-defined as total profit earned by a participant over trials 7-12 of the experiment, denoted $S\pi_i \equiv \sum_{t=7}^{12} \pi_{it}$. Focusing on trials 7-12 here is designed to reduce the 'learning by doing' effects on response variables. Each observation of this new profit variable is clearly influenced by the set of 6 randomly-selected rows faced by the individual. The effect is to shift the empirical distribution of the profit variable away from the lower bound of zero, and also to increase the number of feasible values. The results show that this aggregation largely attenuates non-Normality (see below). The effect on equation (1) is to remove trial as a determinant of profit, leaving a pure cross-section of aggregated profit observations to be explained by the eight remaining variables. Our other response variable is analogously re-defined as $SIR_i \equiv \sum_{t=7}^{12} IR_{it}$ where IR_{ij} is the sum of information requests made in trial

t by individual i. Here again, aggregation over trials gives an approximately Normal distribution for this variable.

Consider first a linear version of equation (1), aggregated over trials 7-12. The profit realised by individual i depends on individual characteristics, including incentive parameters, as well as the collection of six rows faced in Trials 7-12. Thus, equation (1) takes the form

$$\begin{aligned} S\pi_i &= \gamma_0 + \gamma_1 r_i^{\pi} + \gamma_2 c_i^{IR} + \gamma_3 age_i + \gamma_4 mf_i + \gamma_5 upg_i \\ &\quad + \sum_{j=2}^{4} \gamma_{6j} nat_{ji} + \gamma_7 gpa_i + \sum_{k=2}^{20} \gamma_{8k} SR_{ik} + \varepsilon_i \quad i = 1,...,94 \quad . \end{aligned} \tag{2}$$

The row effect could, in principle, be non-linear across rows, and this is modelled in (2) by allowing mean shifts in the intercept conditional on the rows faced, with SR_{ik} denoting the number of times row k occurred for individual i in trials 7-12. Since $\sum_{k=1}^{20} SR_{ik} = 6$ for all i, the specification in (2), which includes a general intercept, omits the dummy SR_{i1}, in order to avoid singularity problems in the estimation procedure.

Equation (2) can be estimated, but it is not exactly parsimonious. The row effects require estimation of 19 parameters $\gamma_{8k}, k = 2,...,20$. However, given the structure of the profit function in Table 2, it can be expected, *ceteris paribus*, that higher profits will generally be earned the higher the row faced. Indeed, given the approximately linear way in which the maximum profit increases with Row, a possible simplification involves the hypothesis that the parameter γ_{8k} will vary linearly with row k. This implies a restriction of the form

$$\gamma_{8k} = k\gamma_8 \qquad \text{for } k = 1,..,20 . \tag{3}$$

If this is a statistically acceptable restriction, it allows a major reduction in the number of parameters requiring estimation. Notice that given (3), it is possible to write $\sum_{k=2}^{20} \gamma_{8k} SR_{ik} = \gamma_8 \sum_{k=2}^{20} kSR_{ik}$. Thus, defining a weighted row-count variable CR for individual i as

$$CR_i \equiv \sum_{k=2}^{20} kSR_{ik} \ , \tag{4}$$

equation (2) can be written as

$$S\pi_i = \gamma_0 + \gamma_1 r_i^{\pi} + \gamma_2 c_i^{IR} + \gamma_3 age_i + \gamma_4 mf_i + \gamma_5 upg_i$$
$$+ \sum_{j=2}^{4} \gamma_{6j} nat_{ji} + \gamma_7 gpa_i + \gamma_8 CR_i + \varepsilon_i \quad i = 1,...,94 . \tag{5}$$

Parallel specifications to (2) and (5) may also be applied to the other response variable, SIR_i, though admittedly the motivation for incorporating CR_i is not so clear in this case. It turns out that, for both response variables, restricting the models to one in which aggregate response is *linear in row number* is statistically acceptable; the regressions and testing down procedure that establish this are given in Appendix Three. Using the variable *CR*, a weighted sum of row numbers, with weights equal to the number of times each row occurred in trials 7-12, is a relatively parsimonious form of row control. Given its statistical performance, it is used throughout the subsequent analysis. Table 7 presents notation, definitions and feasible values for variables used in the analysis.

Table 7: Notation, definitions and feasible values for response variables and determinants

Definition	**Notation**	**Range of Values**
Response Variables		
Aggregate profit earned by participant in Trials 7-12	$S\pi_i$	0-3000
Number of information requests made by participant in Trials 7-12	SIR_i	0-30
Determinants		
Weighted sum of profit table row numbers faced by participant in Trials 7-12	CR_i	6-120
Money reward paid to participants per unit of profit earned (pounds)	r_i^{π}	0-0.02
Money cost to participants per information request (pounds)	c_i^{IR}	0.01-0.20
Age of participant (years)	age_i	18-
Sex of Participant	mf_i	Female: $mf = 1$, Male: $mf = 0$
Graduate status of participant	upg_i	Postgraduate: $upg = 1$; Undergraduate: $upg = 0$
Dummy variables for Nationality of participant (British, non-British European, Asian or other)	Nat_{2i} Nat_{3i} Nat_{4i}	British: $nat2=nat3=nat4=0$; European non-British: $nat2=1$, $nat3=nat4=0$; Asian: $nat3=1$, $nat2=nat4=0$; Other: $nat4=1$, $nat2=nat3=0$
Grade point average (%)	gpa_i	0-100

The only further refinement of equation (5) we considered was to allow for possibly differential effects of the four sub-groups, 'G0', 'G1', 'G2' and 'G3' derived from the post-experiment questionnaire. Dummy variables were created to allow for differences in the constant term, the row control variable *CR*, and the incentive variables c^{IR} and r^{π}. For m = 0, 1, 2, 3, let the dummy variable Gm_i = 1 if participant i is a member of sub-group 'Gm'; and 0 otherwise. Then the estimation model for profit, corresponding to equation (5), takes the form

$$\begin{aligned} S\pi_i = \sum_{m=0}^{3}\left(\gamma_{0m}Gm_i\right) + \sum_{m=0}^{3}\left(\gamma_{1m}r_i^{\pi}Gm_i\right) + \sum_{m=0}^{3}\left(\gamma_{2m}c_i^{IR}Gm_i\right) + \gamma_3 age_i + \gamma_4 mf_i \\ + \gamma_5 upg_i + \sum_{j=2}^{4}\gamma_{6j}nat_{ji} + \gamma_7 gpa_i + \sum_{m=0}^{3}\left(\gamma_{8m}CR_i Gm_i\right) + \varepsilon_i \quad i = 1,\ldots,94 \end{aligned} \tag{6}$$

The general intercept is still present in (6) as γ_{00}. The use of multiplicative dummies implies that, for example, the coefficient of r_i^{π} applicable to sub-group 'G0' is γ_{10}; the coefficient of r_i^{π} applicable to sub-group 'G1' is $\gamma_{10} + \gamma_{11}$; the coefficient of r_i^{π} applicable to sub-group 'G2' is $\gamma_{10} + \gamma_{11} + \gamma_{12}$; and the coefficient of r_i^{π} applicable to sub-group 'G3' is $\gamma_{10} + \gamma_{11} + \gamma_{12} + \gamma_{13}$. That is, γ_{1m} for $m \geq 1$ represents the incremental effect over and above that for $\sum_{j=0}^{m-1}\gamma_{1j}$. The usual tests can then reveal, for each affected determinant and for the constant term, whether sub-groups differ amongst themselves, or are homogeneous and therefore can be pooled, saving degrees of freedom in estimation. A similar specification is used in the model for SIR_i. Two versions of equation (6) were estimated: 'unrestricted' and 'no pooling'. The 'no pooling' version assigned the parameters $\gamma_{0m} = \gamma_{1m} = \gamma_{2m} = \gamma_{8m}$, $m = 0,1,2$, and restricted the estimation to the G3 sub-sample of 42 subjects correctly answering all questions in the post-experiment questionnaire.[18] For each version, a multi-stage procedure of testing down was employed, whereby at each stage the single most insignificant variable was omitted, the equation re-estimated, diagnostic tests re-checked, and the equation re-examined for other insignificant variables that might be omitted. For convenience of interpretation, we reproduce the research hypotheses from Section 1 here, using the notation in Table 7.

H_1: As r^{π} increases, *SIR* and $S\pi$ increase.

H_2: As c^{IR} increases, *SIR* and $S\pi$ decrease.

[18] The 'no pooling' version of equation (6) is thus equivalent to estimating equation (5) on the G3 sub-sample.

H_3: As *mf* increases, $S\pi$ decreases.

H_4: As *age* increases, $S\pi$ decreases.

H_5: As *gpa* increases, $S\pi$ increases.

H_6: As *nat3* increases, $S\pi$ decreases.

H_7: As *upg* increases, $S\pi$ increases.

Table 8: Results for Aggregate Information Requests (SIR)

Equation 6 version	Tested Down OLS Regression
'no pooling' (sample size 42)	$SIR = \underset{(0.00)}{10.08} + \underset{(0.00)}{465.1} r^{\pi} - \underset{(0.00)}{41.7} c^{IR} + \underset{(0.01)}{9.32} nat3$ Adj. R^2: 0.35, DW: 2.18 (0.85), F (zero slopes): 8.24 (0.00), Jarque-Bera: 2.12 (0.35) LM Het: 1.22 (0.27), Ramsey's RESET2: 0.14 (0.71)
'unrestricted' (sample size 94)	$SIR = \underset{(0.74)}{1.78} - \underset{(0.04)}{5.72} G2 + \underset{(0.05)}{12.28} G3 + \underset{(0.00)}{338.36}\left(r^{\pi} \times G0\right) - \underset{(0.00)}{31.22}\left(c^{IR} \times G0\right)$ $+ \underset{(0.06)}{0.4}\, age - \underset{(0.00)}{8.94} nat4 + \underset{(0.03)}{0.09}\left(CR \times G1\right) - \underset{(0.06)}{0.18}\left(CR \times G3\right)$ Adj. R^2: 0.27, DW: 2.09 (0.91), F (zero slopes): 5.38 (0.00), Jarque-Bera: 0.34 (0.84) LM Het: 0.75 (0.39), Ramsey's RESET2: 0.01 (0.92)

Probability values, given in parentheses, are one-tailed whenever the research hypothesis is signed.
SIR denotes aggregate information requests made by a subject in trials 7-12 of the experiment; r^{π} denotes money reward (pounds) paid to participants per unit of profit earned; c^{IR} denotes money cost (pounds) to participants per information request; *nat*3 indicates subject is of Asian nationality; *nat*4 indicates subject is of non-European and non-Asian nationality; *age* denotes age of subject; *CR* is a variable controlling for the profit-table rows faced by the subject during the last 6 trials of the experiment; *G*0,…,*G*3 denote increasingly restricted nested sub-groups of the sample formed on the basis of responses to the post-experiment questionnaire (*G*0 is the full sample and *G*3 is the smallest sub-sample).

Discussion of Results for Hypotheses 1 and 2

Table 8 presents the results for the information requests response variable, *SIR*. There are two hypotheses, H_1 and H_2, relevant to this variable. For the smaller sample of G3 subjects, 'no pooling', the two hypotheses are clearly supported, with one-tailed probability values <<1%. When the analysis makes use of the whole sample G0-G3, 'unrestricted', it is clear from the presence and significance of the composite variables $\left(r^{\pi} \times G0\right)$ and $\left(c^{IR} \times G0\right)$ that support for H_1 and H_2 extends to participants who failed to answer some or all of the post-experiment questionnaire. Although when the analysis is extended from G3 to G0-G3 the marginal effects of varying rewards and costs are diluted slightly, this dilution is *not significant*. With regard to demographic variables, the only significant determinant of information requests in the G3 group alone is Asian nationality, which has a positive coefficient. When the entire sample, G0-G3, is used, however, this variable loses its significance. Instead for this larger cohort, age has a slight positive impact and Non-European non Asian nationality has a larger negative impact.[19]

Table 9: Results for Aggregate Profit (Sπ)

Equation 6 version	Tested Down OLS Regression
'no pooling' (sample size 42)	$S\pi = \underset{(0.22)}{-237.3} + \underset{(0.00)}{14856} r^{\pi} - \underset{(0.19)}{533.64} c^{IR} - \underset{(0.03)}{145.78} mf + \underset{(0.00)}{19.19} CR$ Adj. R^2: 0.57, DW: 2.32 (0.95), F (zero slopes): 14.34 (0.00), Jarque-Bera: 2.72 (0.26) LM Het: 0.54 (0.46), Ramsey's RESET2: 0.46 (0.50)
'unrestricted' (sample size 94)	$S\pi = \underset{(0.92)}{22.88} - \underset{(0.13)}{424.27} G1 + \underset{(0.00)}{14202.9}(r^{\pi} \times G3) - \underset{(0.11)}{667.94}(c^{IR} \times G3)$ $- \underset{(0.09)}{74.26} mf - \underset{(0.01)}{168.69} nat3 - \underset{(0.00)}{331.53} nat4$ $+ \underset{(0.00)}{15.03}(CR \times G0) + \underset{(0.01)}{11.64}(CR \times G1) - \underset{(0.01)}{4.88}(CR \times G2)$ Adj. R^2: 0.55, DW: 1.96 (0.80), F (zero slopes): 13.77 (0.00), Jarque-Bera: 0.14 (0.94) LM Het: 2.36 (0.12), Ramsey's RESET2: 0.00 (0.97)

Probability values, given in parentheses, are one-tailed whenever the research hypothesis is signed.
SIR denotes aggregate information requests made by a subject in trials 7-12 of the experiment; r^{π} denotes money reward (pounds) paid to participants per unit of profit earned; c^{IR} denotes money cost (pounds) to participants per information request; *mf* =1 indicates subject is female; *nat*3 =1 indicates subject is of Asian nationality; *G*0,…,*G*3 denote increasingly restricted nested sub-groups of the sample formed on the basis of responses to the post-experiment questionnaire (*G*0 is the full sample and *G*3 is the smallest sub-sample); *CR* is a variable controlling for the profit-table rows faced by the subject during the last 6 trials of the experiment;

The results for the profit response variable are presented in Table 9. All the research hypotheses, H_1 to H_7, are relevant here. For the 'no pooling' analysis of the G3 sub-group, the rate of personal reward per profit point earned, r^{π}, is significantly positively related to earned profit, as per H_1. The incentive coefficient value of 14,856 is equivalent to about 12 extra profit points per trial *for every halfpenny increase in* r^{π}. Thus relative to fixed rewards (FR), performance-related pay (PRR) offers a substantial amount of leverage through payment by results. Moreover, this effect is shown to survive when the saliency of personal money rewards is increased approximately by a factor of

[19] No subjects with a *nat4* coding made it into the G3 group, thus explaining its absence from the 'no pooling' equation.

four from the level in Sprinkle (2000), strengthening claims to external validity. Varying costs to accessing information, as per H_2, for the G3 sub-group produces the predicted negative coefficient. However, with a one-tailed p-value of 19%, the coefficient is not significant. Taken together with the significant finding for G3 of a negative impact on number of information requests in Table 8, this presents an apparent anomaly. One might conjecture that increasing information cost leads rationally to economising on usage, but might also lead to more effort then being expended on making the best possible use of the information that has been accessed; it is possible the latter effect counters the former, and hence explains the observed result.[20]

The results pertaining to profit for the 'unrestricted' model are similar to those for the G3 sub-group, with H_1 strongly supported, and H_2 not supported albeit with a lower p-value of 11%. Note however, that the tested-down 'unrestricted' model for profit includes only two incentives variables, the composites $\left(r^{\pi} \times G3\right)$ and $\left(c^{IR} \times G3\right)$. This means that despite the fact all subjects in G0 vary their demand for information in like manner when incentives are varied, this variation translates into changes in profit performance for *only* the G3 sub-group answering all of the questionnaire correctly. It is possible that this apparent anomaly may be due to the manner in which subjects were coded into groups, according to anticipated profit-earning capability rather than anticipated information use. Thus, for instance, G0 subjects were excluded from higher sub-groups because they failed to answer questions regarding the profit table, not because they failed to understand their money rewards and costs. Thus the G0 group may respond to variation in money rewards and costs, but are unable to translate this response into profits.[21] This explanation is not adequate, however, for the finding of no profit response for the G1 and G2 groups, because these groups correctly answered the profit table questions but not the incentive questions. Another possible explanation, with implications for the design of PRR systems outside the laboratory, is that PRR systems will produce benefits in company profitability only when offered to subjects who, for whatever reason, are in a position to deliver these benefits, otherwise with risk-averse employees PRR will simply cost more on average than FR; see Kaplan & Atkinson (1998, p.679). In this case, PRR is only effective at raising performance when subjects have full understanding. A metric of

[20] No comparison can be made with Sprinkle (2000) here, because in that paper, information cost was not varied in the experimental design.

[21] Note that the finding of a behavioural response to variation in r^{π} and c^{IR} for subjects outside the G3 group would not have been observed if these subjects had been discarded on the basis of the post-experiment questionnaire, whereas if the problem of task-incentive comprehension had been ignored and the data for all 94 subjects had been pooled, then the finding of a *differential* response in profit performance would not have been observed. The approach we have adopted here in the 'unrestricted' model thus provides insights that may otherwise be overlooked, particularly when the experimental environment is complex and comprehension becomes a significant issue.

comprehension/rationality, based upon the proportion of *dominated* column choices selected by PRR subjects, showed substantial improvements in the G3 group compared to the full sample.[22] The proportion of dominated column choices for the G3 group in the present study was similar to the metric reported in Sprinkle (2000), thus suggesting why a significant result was obtained in that paper.

Discussion of Results for Hypotheses 3 to 7

The coefficients of *mf*, *age*, *gpa*, *nat3*, and *upg* all had signs predicted by the research hypotheses, in both the 'no pooling' and the 'unrestricted' models, but not all these factors appear in the tested-down models because not all were significant. The most clear-cut result is for H_3: the effect of gender. Female subjects did significantly less well than males, in both the G3 group and overall in the 'unrestricted' model. The negative effect of substituting a female participant for a male participant is to lower expected profit in trials 7-12 by about between 74 and 146 units of profit, or about 12-24 units of profit per trial; a finding similar to the corresponding result of 20 in Dobbs & Miller (2006). This result offers confirmation of previous results in the literature; that females do less well in problem-solving tasks than males.[23] We found that *age* and *upg* were highly positively correlated in the sample. Both coefficients had the signs predicted by H_4 and H_7, but both were insignificant, a not untypical finding for collinear variables. H_5 predicts *gpa* will have a positive impact on profit performance. Although the coefficient was positive, again it was insignificant and excluded from the tested down equations. It is possible that the G3 coding could well be doing the work for *gpa* in the profit equation. Also *gpa* appears to have some explanatory power in explaining performance in Trials 1-6 (not reported). Higher *gpa* students do better in initial trials, perhaps because they understand the written/oral instructions faster. But after completing several trials, in effect there is 'learning by doing', and the *gpa* effect has disappeared by Trials 7-12. The results for H_6 are ambiguous: *nat3* appears as a significant determinant in the 'unrestricted' equation, with the correct sign, but does not appear in the 'no pooling' G3 equation. Finally, non-European non-Asian nationality, *nat4*, has a significant negative impact on profit performance, just as it did on demand for information in Table 8.[24]

[22] For more details, see Dobbs & Miller (2008).
[23] See Section 1 for references.
[24] There are no *nat4* subjects in the G3 sub-group.

Overall, we find there is a significant role for certain personal characteristics, primarily gender and cultural variables, in explaining behaviour and task performance. For a complex experimental design, the specification of the statistical model can be improved by the inclusion of such variables. In addition our evidence confirms previous findings with regard to gender and culture for the problem-solving task studied here.

4. Conclusions

Though it tends to be under-valued in social sciences research, in the natural sciences, replication is a critical element in the establishment of robust research findings. In many Social Science applications the potential for replication is inherently limited, but in laboratory experimental work this is not the case – replication is always possible – and hence it appears that at least in these applications, the Social Sciences can hope to achieve the level of rigour manifest in good quality scientific research. Using a research design based on Sprinkle (2000), this paper attempted to replicate a previous finding that, compared to fixed rewards, performance-related rewards increase individual demand for valuable information and enhance performance. This issue is of major importance to the design of performance measurement and reward systems based on accounting information, since agency theory suggests that performance-related rewards are necessary to mitigate moral hazard and adverse selection problems in complex organisations.

Our study also examined a more general version of the Sprinkle setting, including the provision of nineteen treatments rather than just two treatments, and so was able to investigate whether the earlier finding survived when incentives are made more salient. Our larger sample also had rather different demographics compared to the original work in Sprinkle (2000); instead of featuring exclusively US sophomore participants, our sample featured final year undergraduate and postgraduate students with a non-US international background (primarily European and Asian students). To explore the specific effects of different demographics, we collected and made use of data on personal characteristics (including GPA, gender, age, nationality). Finally, we undertook a post experiment questionnaire study, and used the data from this as an integral part of the statistical testing procedure. Whilst post-experiment questionnaires have been used in previous work, they have been used exclusively to eliminate subjects who reveal some apparent imperfections in their understanding (make errors in one or more of their answers to the post experiment questions). We take the view that responses to such questionnaires, just as with responses within the experiment, are properly viewed as stochastic variables, and that it is therefore appropriate to explicitly model this within the overall analysis of the experiment as a whole. It is rare (non-existent, to our knowledge) for the data collected in post experiment questionnaires to be formally incorporated into the statistical analysis. In our view, if appropriately designed, the data that can be gathered from such a source can profitably be used as an integral part of the analysis via a general-to-specific testing down methodology. The present study illustrates the value of this approach

Our tests confirm the result originally found in Sprinkle (2000), but for a much wider range of incentives: when information is valuable for enhancing profit, but is also costly, its effective use by individuals can be increased if individuals are given profit-related incentives. This lends support to the predictions of agency theory as it relates to accounting issues. However, we also show that the finding must be qualified, for although all subjects responded to incentives similarly, only those with a better understanding of the task and incentives could translate this into significantly better performance on the task. It may be that in Sprinkle (2000) this qualification was obscured because comprehension of the task and incentives was much less of an issue in that study than in the present one. Nevertheless, it is an important qualification, for it confirms an aspect of agency theory dealing with executive compensation; namely, that the benefit to a company of providing performance-related rewards is not universal and depends, inter alia, on the ability of an individual to influence profit. Interestingly, we also find, even for subjects with a better understanding of the task and incentives, that increasing the cost of information leads to significantly less accessing of information and some decrease in profit performance, but that the weakening of performance is not statistically significant. This issue, which was not observed in Sprinkle (2000), where the cost of accessing information was a fixed datum that never varied, requires further investigation. Finally, we have shown performance-related rewards have the predicted effect on demand for information and task performance for a sample with different demographics than in Sprinkle (2000). For this particular problem-solving task we also confirm findings in other studies that gender and culture play a role in determining performance independently of the structure of rewards.

With regard to the relevance of the study to accounting practice, we do not wish to over-claim the importance of the investigations presented here. The laboratory approach used involved structuring a controlled business-like environment and designing explicit information and reward structures for this environment. Participants then played an experimental ‘game’ and earned real money rewards as a consequence of their decisions. The advantage of the approach is that the experiment can be designed to explicitly address the issue under investigation, while at the same time removing much of the confounding influence of extraneous variables in naturally-occurring settings. The experimental approach has its drawbacks, in part because the experiment environment is intrinsically somewhat artificial, and the amount of time that one can ask participants to commit is usually relatively short; in the present study up to about 90 minutes. Hence it is not usual to claim it simulates a naturally-occurring setting. However, the approach brings with it much of the power of scientific method, and is perhaps best suited for testing the simplified models typically propounded in theoretical work. Since a theoretical model can only be considered general if it applies to all its proper special cases, analysis of data generated from simple laboratory experiments can therefore be said to offer legitimate tests of model generality, regardless of differences between experimental and naturally-occurring environments; see Plott (1991), Roth (1991), Starmer (1999) and Smith (2002) for further discussion.

Finally, there are other issues that could be further investigated: the significance of a learning effect as participants gain experience of the decision-making task; the nature of 'rules of thumb' followed by participants when taking decisions; the impact of risk attitudes on behaviour; and more technical issues such as the adequacy of controls for the effects of profit table row on performance, interactions between determinants, and the functional form of the data generating process. Additional experiments could also be undertaken to examine how variations in the structure of profit-related money rewards, other than saliency, affect the use of information in a complex decision-making environment. For instance, the use of 20-20 hindsight in performance evaluation could be investigated in much the same framework as the present study.

References

Arkes, H. R., R. M. Dawes and C. Christensen (1986), Factors Influencing the Use of a Decision Rule in a Probabilistic Task, *Organizational Behavior and Human Decision Processes*, Vol.37, February, pp.93-100.

Arrow, K. J. (1970), The Theory of Risk Aversion, *Essays in the Theory of Risk-Bearing*, North-Holland Publishing Co., Amsterdam.

Ashton, R. H. (1990), Pressure and Performance in Accounting Decision Settings: Paradoxical Effects of Incentives, Feedback, and Justification, *Journal of Accounting Research*, Vol.28, Supplement, pp.148-180.

Baiman, S. (1982), Agency Research in Managerial Accounting: A Survey, *Journal of Accounting Literature*, Vol. 1, pp.154-213.

Baiman, S. (1990), Agency Research in Managerial Accounting: A Second Look, *Accounting, Organizations and Society, Vol. 15(4), pp.341-371.*

Baiman, S. & J. S. Demski (1980), Economically Optimal Performance Evaluation and Control Systems, *Journal of Accounting Research, Supplement, pp.184-220.*

Berg, J. E., L. A. Daley, J. W. Dickhaut and J. R. O'Brien (1986), Controlling Preferences for Lotteries on Units of Experimental Exchange, *The Quarterly Journal of Economics*, Vol. 101(2), pp.281-306.

Binmore, K. (1994), *Playing Fair*, Cambridge, Mass., MIT Press.

Bonetti, S. (1998), Experimental Economics and Deception, *Journal of Economic Psychology*, Vol. 19(3), June, pp.377-395.

Bonner, S. E. (1999) Judgement and Decision-Making Research in Accounting, *Accounting Horizons*, Vol. 13, pp.385-398.

Bradley, J. V. (1978), Robustness? *British Journal of Mathematical and Statistical Psychology*, Vol. 31, pp.144-152.

Chow, C. W., K. Haddad & G. Totterdale (1990), The Efficiency and Effectiveness of Commonly Used Cost Variance Investigation Rules, *The British Accounting Review*, Vol. 22(1), March, pp.27-40.

Ciancanelli, P., S. Gallhofer, C. Humphrey & L. Kirkham (1990), Gender and Accountancy: Some Evidence from the UK, *Critical Perspectives on Accounting, Vol. 1, pp.117-144.*

Clinch, J. J. & H. J. Keselman (1982), Parametric Alternatives to the Analysis of Variance, *Journal of Educational Statistics*, Vol. 7, pp.207-214.

Davis, D. D. & C. A. Holt (1993), *Experimental Economics*, Princeton University Press, Princeton, New Jersey.

Davis S., F. T. DeZoort and L. S. Kopp (2006), The Effect of Obedience Pressure and Perceived Responsibility on Management Accountants' Creation of Budgetary Slack, *Behavioral Research in Accounting*, Vol. 18, pp.19-35.

Deaux, K. (1979), Self-Evaluations of Male and Female Managers, *Sex Roles, Vol. 5(5), pp.571-580.*

Dermer, J. D. (1973), Cognitive Characteristics and the Perceived Importance of Information, *The Accounting Review, Vol. 48(3), pp.511-519.*

Dobbs, I. M. and A. D. Miller (2006), *The Impact of Financial Incentives on Decision Making*, The Institute of Chartered Accountants of Scotland, Edinburgh.

Dobbs, I. M. and A. D. Miller (2008), *The Impact of Financial Incentives on Decision Making: Further Evidence*, The Institute of Chartered Accountants of Scotland, Edinburgh.

Drake, A. R., S. F. Haka and S. P. Ravenscroft (1999), Cost System and Incentive Structure Effects on Innovation, Efficiency, and Profitability in Teams, *The Accounting Review*, Vol.74, pp.323-345.

Feltham, G. A. (1972), *Information Evaluation*, American Accounting Association, Sarasota, Florida.

Fisher, J. G., L. A. Maines, S. A. Peffer and G. B. Sprinkle (2002), Using Budgets for Performance Evaluation: Effects of Resource Allocation and Horizontal Information Asymmetry on Budget Proposals, Budget Slack, and Performance, *The Accounting Review*, Vol. 77(4), pp.847-865.

Fisher, J. G., S. A. Peffer and G. B. Sprinkle (2003), Budget-Based Contracts, Budget Levels and Group Performance, *Journal of Management Accounting Research*, Vol. 15, pp.51-74.

Glass, G. V., P. D. Peckham & J. R. Sanders (1972), Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analysis of Variance and Covariance, *Review of Educational Research*, Vol. 42, pp.237-288.

Guala, F. (2005), *The Methodology of Experimental Economics*, Cambridge University Press, New York.

Hambur, S., K. Rowe & L. T. Luc (2002), Graduate Skills Assessment: Stage One Validity Study, *Commonwealth of Australia.*

Hirshleifer, J. & J. G. Riley (1992), *The Analytics of Uncertainty and Information*, Cambridge University Press, Cambridge, UK.

Hogarth, R. M., B. J. Gibbs, C. R. McKenzie and M. A. Marquis (1991), Learning from Feedback: Exactingness and Incentives, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol.17, July, pp.734-752.

Kaplan, R. S. & A. A. Atkinson (1998), *Advanced Management Accounting, 3rd edition,* Prentice-Hall, New Jersey, USA.

Kuncel, N. R., Hezlett, S. A. & D. S. Ones (2004), Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All? *Journal of Personality and Social Psychology, Vol. 86(1), pp.148-161.*

Lundeberg, M. A., P. W. Fox & J. Punćochaŕ (1994), Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgements, *Journal of Educational Psychology, Vol. 86(1), pp.114-121.*

Maddala, G.S. (2001), *Introduction to Econometrics*, 3rd ed., Wiley, Chichester, England.

Plott, C. R. (1982), Industrial Organization Theory and Experimental Economics, *Journal of Economic Literature*, Vol. 20, pp.1485-1527.

Plott, C. R. (1991), Will Economics Become and Experimental Science? *Southern Economic Journal, Vol. 57(4), April, pp.901-919.*

Prasnikar, V. (2002), Do Binary Lottery Payoffs Measure Deviations from Expected Utility? An Experimental Study, *Discussion Paper, Graduate School of Industrial Administration, Carnegie Mellon University.*

Rabin, M. (2000), Risk Aversion and Expected Utility Theory: A Calibration Theorem, *Econometrica*, Vol. 68(5), pp.1281-1292.

Riecken, H. W. (1962), A Program for Research on Experiments in Social Psychology, *Washburne, N F (ed), Decisions, Values and Groups, Volume Two: Proceedings of a Conference Held at the University of New Mexico*, Symposium Publications Division, Pergamon Press, Oxford, pp.25-41.

Ravenscroft, S. and S. Haka (1996), Incentive Plans and Opportunities for Information Sharing, *Behavioral Research in Accounting*, Vol. 8, pp.114-133.

Research and Library Services (2001), Selected Current Research in Gender and Educational Attainment, *Research Paper 06/01*, Northern Ireland Assembly.

Rosenthal, R. (1963), On the Social Psychology of the Psychological Experiment: The Experimenter's Hypothesis as Unintended Determinant of Experimental Results, *American Scientist*, Vol. 51, pp.268-283.

Roth, A. E. (1991), Game Theory as a Part of Empirical Economics, *The Economic Journal, Vol. 101, January, pp.107-114.*

Rubinstein, A. (2001), A Theorist's View of Experiments, *European Economic Review*, Vol. 45, pp.615-628.

Selten, R., A. Sadrieh and K. Abbink (1999), Money Does Not Induce Risk Neutral Behavior, But Binary Lotteries Do Even Worse, *Theory and Decision*, Vol. 46, pp.211-249.

Siegel, S. & D. A. Goldstein (1959) Decision-Making Behavior in a Two-Choice Uncertain Outcome Situation, *Journal of Experimental Psychology, 57, 37-42.*

Smith, V. L. (1994), Economics in the Laboratory, *Journal of Economic Perspectives*, Vol. 8, pp.113-131.

Smith, V. L. (2002), Method in Experiment: Rhetoric and Reality, *Experimental Economics, Vol. 5, pp.91-110.*

Sprinkle, G. B. (2000), The Effect of Incentive Contracts on Learning and Performance, *The Accounting Review*, Vol. 75, pp.299-326.

Sprinkle G. B. (2003), Perspectives on Experimental Research in Managerial Accounting, *Accounting, Organizations and Society*, Vol. 28, pp.287-318.

Starmer, C. (1999), Experiments in Economics: Should we Trust the Dismal Scientists in White Coats? *Journal of Economic Methodology, Vol. 6(1), pp.1-30.*

Starmer, C. & R. Sugden, Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation (1991), *American Economic Review*, Vol. 81, pp.971-978.

Tan, W. Y. (1982), Sampling Distributions and Robustness of *t*, *F* and Variance-Ratio in Two Samples and ANOVA Models with respect to Departure from Normality, *Communications in Statistics – Theory and Methods*, Vol. 11, pp.486-511.

Tuttle, B. and F. G. Burton (1999), The Effects of a Modest Incentive on Information Overload in an Investment Analysis Task, *Accounting, Organizations and Society*, Vol.24, pp.673-687.

Waller, W. S. & C. W. Chow (1985) The Self-Selection and Effort Effects of Standard-Based Employment Contracts: A Framework and Some Empirical Evidence, *The Accounting Review*, Vol. 60, pp.458-476.

Waller, W. S. (1995) Decision-Making Research in Managerial Accounting: Return to Behavioral-Economics Foundations, Ashton, R. H. & A. H. Ashton (Eds.), *Judgement and Decision Making Research in Accounting and Auditing*, New York, NY, Cambridge University Press, pp.29-54.

Wilcox, R. R. (1995), ANOVA: A Paradigm for Low Power and Misleading Measures of Effect Size, *Review of Educational Research*, Vol. 65, pp.51-77.

Wilcox, R. R. (1998), How many Discoveries have been Lost by Ignoring Modern Statistical Methods? *American Psychologist*, Vol. 53, pp.300-314.

Appendix 1: Instructions for Participants with Profit-Related Rewards

Materials

A copy of instructions (this document), a table of numbers, a pen, and paper are provided. Check you have these items now. If anything is missing please raise your hand and I will attend to you. These materials should be returned to the instructor supervising in the computer cluster at the end of the experiment.

Rules of Communication

Our purpose in conducting this experiment is to gain information about *individual* behaviour rather than *group* behaviour. For this reason, we must insist on the following rules concerning communication:

(a) No communication - there must be no verbal or other forms of communication between those taking part until everyone has completed the experiment,

(b) No questions – the only clarifications you can seek are those concerning the use of software in the PC room (if you have difficulty using the software, raise your hand and wait for assistance from the instructor).

Introduction

You will be taking part in a timed computerised experiment designed to investigate the nature of individual decision-making under conditions of uncertainty. You will be paid for your participation. Your earnings will depend upon how you perform the experimental task. Earnings will be calculated and reported to you today, privately – they are then paid in cash, privately, in a few days time. I will first explain the experimental task and then how you earn money from performing this task.

The Experimental Task

Please look at your profit table. The rows are labelled 1 to 20 in the left hand column. The columns are labelled 1 to 20 in the top row. In essence you will be playing a series of 'games' against the computer in which the computer first randomly picks a row, which you are unable to observe; and then you choose a column. The chosen row and column will determine, where they cross in the profit table, the profit earned.

To illustrate how the game works, suppose one of you plays 'the computer' [SELECT SOMEONE TO PLAY THIS ROLE – NAMED "X" BELOW]. X should write down a row number on a piece of paper, without revealing it to the group, and then pass the piece of paper to me. Now, consider what happens when you make a choice of column. [ASK A DIFFERENT RESPONDENT "Y" FOR A CHOICE OF COLUMN: THIS IS THEN ANNOUNCED SO EVERYONE CAN HEAR THE CHOICE]. The profit result is determined by the intersection of the row and column. In fact the row chosen was …. ("X" CONFIRMS THE CHOICE OF ROW). Hence the profit is ….(EXPLAIN THE RESULT VIA AN OVERHEAD).

The experiment is organised as follows:-

(a) There are 12 trials, for each of which there are 5 decision periods, altogether making 12×5=60 column decisions for you to take. You identify your column choice each time by inputting your chosen column's number at your computer terminal when prompted. The number of the column you want to choose can be found in the top row of the profit table

(b) For each trial, the computer will select a row, which is then FIXED for all periods in the trial.

(c) You then choose a column in period 1 of this trial.

(d) The computer will ask you whether or not you wish to learn the profit earned by your choice in (c) above. If you agree, it will be reported immediately (but note there is a cost associated with this - see below). If you do not, the profit earned will not be reported until the end of the trial.

(e) The process in (c) and (d) is then repeated for periods 2,..,5 of the trial.

(f) At the end of 5 periods, the profits earned in each period are revealed to you without cost (see information costs below), and your money earnings are calculated and also revealed.

(g) The process now starts again with a new trial – the computer again chooses a row at random which is then fixed during the trial, and you repeat the process (c)-(f), (and so on, until all trials have been completed)

In taking part in the experiment, we would like you to try to earn as much profit as you can.

Notice, in the profit table that, since the chosen row will not be known to you when you select your column, you will generally not know for certain the profit that will be earned by your column choice. However, you *do* know that the row is fixed for the 5 periods of each trial, so that whenever you choose a column, not only does your choice determine profit, jointly with the selected row, but if you make an information request, it also provides information about the unobserved row being used in the current trial. Look again at the profit table. ["Y" who previously chose the column] chose column … If an information request is made, it would be discovered that profit from the decision was …. And one can reason that the computer has selected a row between row… and row… [ILLUSTRATE THIS ON OVERHEAD] With this knowledge, one may (or may not) wish to change the choice of column in the next period (and so on).

Additionally, the following information is available about how the computer has been programmed:

- The rows 1 to 20 are all equally likely to be randomly selected by the computer.

- The rows selected for you by the computer in a given trial will generally vary randomly from trial to trial, and will also differ across other individuals taking part in the experiment.

- The selection of a row by the computer in each trial is not affected by your choices of columns in the current or previous trials.

Money Earnings

Your earnings are (i) a fixed proportion of profit earned, but are also (ii) negatively related to the number of times you request profit information *during* a trial. Your money reward per unit of profit earned, and the money loss per information request will be notified to you at the start of the experiment.

You have 180 seconds to complete the five column choices expected in each trial. If you run out of time before completing the five column choices, you will only be paid for those you actually complete. The clock will start anew at the beginning of each trial. Whether or not you requested information on earned profits *during* each trial, you will also be informed (at no extra cost) of the total profits earned at the end of each trial

We recommend that, before you begin the experiment, you spend some time examining the structure of the profit table. Do not be influenced by the pace at which other respondents work, whether slower or faster than you. Your money rewards will be based solely on *your* performance of the task and no one else's. We have allocated plenty of time, an hour, for you to complete the entire experiment.

Payments & Completion of the Experiment

When you have finished the experiment, please raise your hand and an instructor will attend to the final procedures, including the issuing of a short questionnaire which you should complete immediately and hand to the instructor. A summary of your earnings appears on the computer screen at the end of the experiment. Arrangements for payment of your earnings will be emailed to you in the next few days.

Post-Experiment Briefing

This experiment is funded by the Institute of Chartered Accountants of Scotland, who will publish a report on the findings of the experiment in due course. All published data collected will, of course, be anonymous. In order to preserve the integrity of the ongoing experiment, we are unfortunately unable to offer any feedback on its nature or your role in it until the whole experiment is complete. However, we will email every participant in the experiment when the report is available to let them know where they can obtain a copy

Appendix 2: Post-Experiment Questionnaire

Project No. RES0371 07209

(Sponsored by the Institute of Chartered Accountants of Scotland)

Full Name (Print)……………………………………………….

1. Suppose, for a given trial, the row was fixed at Row 15. Suppose you selected column 18 in the first period. What profit would you earn for the firm in that period?

   Your answer ………………………………………..

2. Suppose you begin a new trial. You do not know the row selected by the computer. If in the first period you choose column 6, request profit information, and learn that profit was zero, what do you conclude about the row number selected by the computer?

   Your answer ………………………………………..

3. Suppose, after a choice of column, you earned 0 for the firm. How much do you earn for *yourself*?

   Your answer ………………………………………..

4. Suppose, after a choice of column, you earned 80 for the firm. How much do you earn for *yourself*?

   Your answer ………………………………………..

5. Suppose, after a choice of column, you make an information request. How much does it cost you?

   Your answer ………………………………………..

Appendix 3: Validating use of the weighted row count variable *CR*

The row count variable was used as a control for row effects. In principle, the effect of row on performance might be non-linear. An initial investigation examined whether moving from the model using individual row dummies in equation (2) to the row count variable model in equation (5) was statistically acceptable. The total number of restrictions involved in moving from equation (2) to (5) is in fact 18. Conducting the usual F-test of these restrictions shows that the simplification to (5) is statistically acceptable. That is, running the regression

$$S\pi_i = \gamma_0 + \gamma_1 r_i^{\pi} + \gamma_2 c_i^{IR} + \gamma_3 age_i + \gamma_4 mf_i + \gamma_5 upg_i + \sum_{j=2}^{4} \gamma_{6j} nat_{ji} + \gamma_7 gpa_i + \sum_{k=2}^{20} \gamma_{8k} SR_{ik} + \varepsilon_i \quad i = 1,...,94$$

gives:

N-Obs 94 N-parameters 29
Sum of squared residuals = .463836E+07
Adjusted R-squared = 0.47

whilst running the regression

$$S\pi_i = \gamma_0 + \gamma_1 r_i^{\pi} + \gamma_2 c_i^{IR} + \gamma_3 age_i + \gamma_4 mf_i + \gamma_5 upg_i + \sum_{j=2}^{4} \gamma_{6j} nat_{ji} + \gamma_7 gpa_i + \gamma_8 CR_i + \varepsilon_i \quad i = 1,...,94$$

yields:

N-obs 94 N-parameters 11
Sum of squared residuals = .581437E+07
Adjusted R-squared = 0.48

The F-statistic for the 18 restrictions is

$$F = \frac{(SSER - SSEU)/J}{SSEU/(n-K)} = \frac{(0.581437E7 - 0.463836E7)/18}{0.463836E7/65} = 0.916 < 1$$

which is clearly not significant.

A parallel analysis for SIR_i gives:

$$SIR_i = \gamma_0 + \gamma_1 r_i^{\pi} + \gamma_2 c_i^{IR} + \gamma_3 age_i + \gamma_4 mf_i + \gamma_5 upg_i + \sum_{j=2}^{4} \gamma_{6j} nat_{ji} + \gamma_7 gpa_i + \sum_{k=2}^{20} \gamma_{8k} SR_{ik} + \varepsilon_i \quad i = 1,...,94$$

where:

N-obs 94 N-parameters 29
Sum of squared residuals = 3053.57
Adjusted R-squared = 0.21

whilst running the regression

$$SIR_i = \gamma_0 + \gamma_1 r_i^{\pi} + \gamma_2 c_i^{IR} + \gamma_3 age_i + \gamma_4 mf_i + \gamma_5 upg_i$$
$$+ \sum_{j=2}^{4} \gamma_{6j} nat_{ji} + \gamma_7 gpa_i + \gamma_8 CR_i + \varepsilon_i \quad i = 1,...,94$$

yields:

N-obs 94 N-parameters 11
Sum of squared residuals = 3835.27
Adjusted R-squared = 0.22

The F-statistic for the 18 restrictions is

$$F = \frac{(SSER - SSEU)/J}{SSEU/(n-K)} = \frac{(3835.27 - 3053.57)/18}{3053.57/65} = 0.924 < 1$$

which again is clearly not significant.

Thus the restrictions are statistically acceptable for both dependent variables, $S\pi$ and *SIR*; this motivates the use of the weighted average row count variable *CR* in the analysis presented in the text.